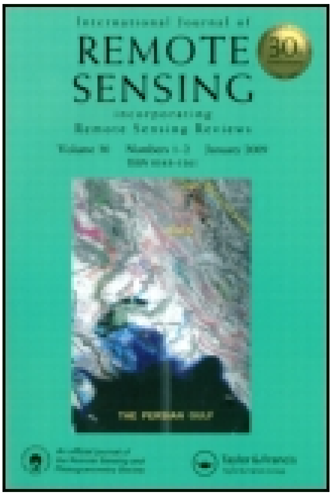

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

General approach to the formulation and solution of the multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types

V.S. Kostsov[a]

[a] Department of Atmospheric Physics, Faculty of Physics, St Petersburg State University, 198504 St Petersburg, Russia

Published online: 12 Jun 2015.

To cite this article: V.S. Kostsov (2015) General approach to the formulation and solution of the multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types, International Journal of Remote Sensing, 36:11, 2973-3004, DOI: 10.1080/01431161.2015.1054961

To link to this article: http://dx.doi.org/10.1080/01431161.2015.1054961

International Journal of Remote Sensing, 2015
Vol. 36, No. 11, 2973–3004, http://dx.doi.org/10.1080/01431161.2015.1054961

General approach to the formulation and solution of the multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types

V.S. Kostsov *

Department of Atmospheric Physics, Faculty of Physics, St Petersburg State University, 198504 St Petersburg, Russia

(*Received 12 December 2014; accepted 5 April 2015*)

A general approach to the formulation and solution of the multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types is considered, which is based on the concept of the presentation of constraints as virtual measurements of finite accuracy. The advantages of the approach are (1) the possibility to account for all kinds of available information (remote and *in situ* measurements, physical constraints, model predictions); (2) the unified description of different measurements, *a priori* information and constraints in the retrieval algorithm; (3) the possibility to use measurements and *a priori* information of different types in any combination and to assess individual contributions to information content. The approach can be considered as a convenient tool for implementation of different synergistic remote-sensing schemes and for utilization of maximum available information on sought parameters for the purpose of increasing the accuracy of atmospheric remote sensing and providing self-consistency of sets of retrieved parameters. The known formulae for degrees of freedom for signal, averaging kernels, error components, and parameters describing spatial resolution are generalized to the case of multiple sought variables, measurements, and constraints, and the peculiar features of such generalization are discussed. The example is given of the application of the approach to the interpretation of the middle atmosphere limb infrared radiance data obtained in CRISTA (Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere) experiments with special emphasis on the constraints describing coupling of the parameters due to physical processes.

1. Introduction

A large number of inverse problems of atmospheric remote sensing are the so-called ill-posed (or improperly posed) problems. Traditionally, in order to obtain a solution of an ill-posed problem, a kind of *a priori* information should be added to the inversion algorithm. The number of papers that consider the information aspects, the error analysis aspects, and many other questions relevant to ill-posed inverse problems is enormous; therefore, we refer to the well-known book by Rodgers (2000) in which all above-mentioned aspects and questions are summarized and discussed in detail.

The classic inverse problem of the remote sensing of atmospheric temperature or composition can be briefly described as follows:

- There is an unknown spatial distribution (profile) of a parameter of atmospheric state to be retrieved.

*Email: vlad@troll.phys.spbu.ru

- There is an experiment measuring a radiation quantity, which is influenced by spatial distribution of a sought parameter.
- There is *a priori* information on a sought parameter, which should be incorporated into the algorithm of the radiative transfer operator inversion in order to obtain the solution.

So, in the simplest case, we have three major components of an inverse problem: an unknown profile of one parameter, an experiment, and *a priori* information. In general case, there may be several unknown parameters, several experiments that use different instruments, and a number of constraints that determine *a priori* information:

- The unknown parameters of primary interest are called the target parameters and the parameters that influence the retrieval and should be preset or controlled in one or another way are called the interfering parameters.
- Combining several different experiments or, in other words, application of synergistic schemes may be very useful for increasing the retrieval accuracy or expanding the spatial range of retrieval of target and interfering parameters.
- When multi-parameter problems are considered, often it is necessary to apply constraints in the form of physical links between different parameters in order to obtain self-consistent retrieval results.

The examples of the multi-parameter inverse problems solved and the inverse problems based on combination of measurements of different types (synergistic methods) are so numerous that any reference list would be obviously incomplete. Therefore, we limit our review to papers that focus on theoretical studies or some kind of generalizations of different aspects of solving inverse problems. First of all, the pioneering paper by Rodgers (1976) should be mentioned in which the virtual measurement concept has been introduced and it has been shown that many of the apparently different methods of solving inverse problems found in the literature are particular cases of the same general method. Rodgers (1998) has shown that the statistical and information theory concepts of information content and degrees of freedom for signal can be effectively used in a straightforward way for optimization of experimental schemes. Practically all aspects of the remote-sensing inverse problems have been presented and analysed in detail in the book (Rodgers 2000) mentioned above. Rodgers and Connor (2003) developed the methods required for intercomparison of measurements made by remote sounders with differing characteristics of the observing systems, particularly their averaging kernels and error covariances, and these methods could be applicable to any kind of retrieval algorithm, not only to optimal estimation. Koner and Drummond (2008) presented a comparison of regularization techniques aimed at maximizing the information content for atmospheric trace gas retrievals and considered four different so-called stabilizers: the identity matrix, the first derivative operator, the second derivative operator, and the *a priori* covariance matrix. Kulawik et al. (2006) introduced a method for developing a constraint matrix based on altitude-varying combinations of zeroth-, first-, and second-order derivatives of the trace gas profile. Doicu et al. (2007) developed an algorithm that combined Tikhonov regularization and the iteratively regularized Gauss–Newton method and was devoted to the solution of multi-parameter inverse problems with simple bounds on the variables. Aires (2011) presented the theoretical study in which various forms of synergy for remote-sensing applications were identified (additive, unmixing, indirect, or denoising synergies) and the methodology was developed to measure these different synergies. The comparison of different retrieval methods has been made by Steck and Clarmann (2001) and it has been shown that an optimal estimation formalism can be used in a highly flexible way. Clarmann, Grabowski, and Kiefer (2001) investigated the problem of accounting for the uncertainty of model parameters in the retrieval algorithm and proposed to use the generalized covariance matrix, which includes, besides measurement noise, also the mapping of parameter errors to the measurement space and leads to a more realistic solution of the inverse problem. A specialized theoretical study of the widely used quantity 'smoothing error' has been done in the recent paper by Clarmann (2014).

Despite the considerable contributions to the theory and practice of inversion methods of atmospheric remote sensing presented in the mentioned papers and many other publications, the idea of Rodgers that many of the apparently different methods of solving inverse problems are particular cases of the same general method did not get the deserved further development. This fact was the main motivation for writing the present article. To date not much attention has been paid to the characterization and analysis of simultaneous use of multiple measurements, multiple constraints and *a priori* information of different types in retrieval algorithms with many unknown parameters, and the general method would be a very convenient tool for doing it. In the present article, the general approach to the formulation and solution of multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types is described based on the concept of the presentation of constraints as virtual measurements of finite accuracy, as proposed by Rodgers (1976). One more circumstance should be mentioned: at present, while reporting results of a solution of an inverse problem, it has become a *de facto* standard to provide a set of additional parameters and characteristics including averaging kernels, degrees of freedom for signal, error components, and parameters describing spatial resolution. In the present article, no new methods are proposed, but an attempt is made to generalize the known formulae to the case of multiple sought parameters, measurements, and constraints, and to describe and discuss the peculiar features of such generalization relevant to estimation of errors and spatial resolution of retrieval results. Besides, an example is given of the application of the general approach to the interpretation of the middle atmosphere limb infrared radiance data obtained in the CRISTA (Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere) experiments (1994 and 1997). As a result of this application, for the first time, the global distributions of vibrational temperatures of four most abundant isotopic variants of the CO_2 molecules were obtained experimentally.

2. Main formulae

2.1. Traditional approach

The traditional approach to the formulation and solution of ill-posed inverse problems of atmospheric remote sensing consists of the following main steps:

- linearization of the radiative transfer equation with respect to a variation of spatial distribution (profile) of one target parameter; the interfering parameters are assumed to be known with certain accuracy;
- determining the cost function with the regularization term;

- deriving the solution;
- estimating the error, the spatial (vertical) resolution of the solution, and the number of independent pieces of information obtained.

Let us consider these steps. We present the linearized radiative transfer equation in the vector–matrix form:

$$\boldsymbol{y} = \boldsymbol{y}_0 + \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0), \tag{1}$$

where $\boldsymbol{y}$ is a vector that is composed of the values of the radiation quantity, for example the radiation intensity or brightness temperature, measured at different frequencies and observation angles; $\boldsymbol{x}$ is the vector composed of the values of the sought parameter (target parameter) at different altitudes (at different coordinates in general case); index '0' denotes reference values for linearization; **K** is the so-called kernel matrix or linearized radiative transfer operator. The interfering parameters are not retrieved. The influence of uncertainty of the values of interfering parameters on the retrieval results is estimated separately.

Since the inverse operator for **K** does not exist in the majority of cases (the problem is ill-posed in the classic sense), solving Equation (1) means searching the minimum of the cost function f_c, which describes the least-square method with the regularization term:

$$f_\text{c} = (\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0))^\text{T}\mathbf{S}^{-1}(\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0)) + (\boldsymbol{x} - \boldsymbol{x}_0)^\text{T}\mathbf{H}(\boldsymbol{x} - \boldsymbol{x}_0), \tag{2}$$

where **S** is the covariance matrix of measurement errors, superscript 'T' denotes transposition, and **H** is the regularization matrix. The most widely used types of regularization are the optimal estimation ($\mathbf{H} = \mathbf{D}^{-1}$, where **D** is the covariance matrix of the sought parameter) and Tikhonov constraint ($\mathbf{H} = \alpha\mathbf{L}^\text{T}\mathbf{L}$, where **L** is the differential operator and α is a parameter that controls smoothness of the solution and can be determined in practical problems by different methods). It is very important to note that $\boldsymbol{x}_\mathbf{0}$ in Equation (2) should be the statistical mean of the sought parameter if optimal estimation is used. It should also be noted that Equation (2) corresponds to application of the least-square method to the following system of matrix equations:

$$\begin{cases} \boldsymbol{y} - \boldsymbol{y}_0 = \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0), \\ \boldsymbol{x}_0 = \boldsymbol{x}, \end{cases} \tag{3}$$

in case of optimal estimation and

$$\begin{cases} \boldsymbol{y} - \boldsymbol{y}_0 = \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0), \\ \mathbf{L}\boldsymbol{x}_0 = \mathbf{L}\boldsymbol{x}, \end{cases} \tag{4}$$

in case of Tikhonov regularization.

In order to find the minimum of the cost function (2), we solve the following equation with respect to $\boldsymbol{x}$:

$$\frac{\partial f_\text{c}}{\partial \boldsymbol{x}} = 0. \tag{5}$$

The solution of Equation (5) and the final solution of the inverse problem will be

$$\boldsymbol{x} = \boldsymbol{x}_0 + \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-\mathbf{1}}\mathbf{K} + \mathbf{H}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}(\boldsymbol{y} - \boldsymbol{y}_0). \tag{6}$$

The error matrix of the solution (the Fisher information matrix) is described by the expression:

$$\mathbf{F} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1}. \tag{7}$$

Often, the error matrix is decomposed into the following sum:

$$\begin{aligned}\mathbf{F} = {} & \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1}\mathbf{H}\left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1} \\ & + \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1},\end{aligned} \tag{7a}$$

where the terms are the so-called smoothing error and retrieval noise matrices, correspondingly (Rodgers 2000).

For the estimation of the spatial resolution of the solution and of the number of independent pieces of information obtained, the formalism of the so-called averaging kernels is often used. If we replace $(\boldsymbol{y} - \boldsymbol{y}_0)$ in Equation (6) with $\mathbf{K}(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0)$, where $\boldsymbol{x}_{\mathrm{t}}$ is the true value of sought vector, we will obtain

$$\boldsymbol{x} = \boldsymbol{x}_0 + \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K}(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0), \tag{8}$$

and

$$\mathbf{A} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} + \mathbf{H}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{K} \tag{9}$$

will be the averaging kernel matrix (AKM); the trace of AKM is equal to the number of independent pieces of information in the solution $\boldsymbol{x}$ delivered by measurement $\boldsymbol{y}$, which is called the degrees of freedom for signal (DOFS; Rodgers 2000). ‘Pieces of information’ should be clarified. For simplicity, let us consider the extreme case when **H** is negligibly small with respect to the first term in parentheses and the operator **K** is expressed by a square matrix which is equal to the identity matrix. Then matrix **A** will be the identity matrix and its trace will be equal to the dimension of the vector of unknowns. Evidently, the dimension of the vector of unknowns determines the maximum number of what is called ‘independent pieces of information’. In an ordinary case, the trace of matrix **A** will always be less than the dimension of the vector of unknowns. If in an ordinary case any diagonal element of matrix **A** is close to unity, then the corresponding element of the vector of unknowns can be considered as the ‘piece of information extracted independently’. And finally, if the diagonal elements of matrix **A** are far from unity, a linear combination of the elements of the vector of unknowns, for which the sum of corresponding diagonal elements of matrix **A** is close to unity, can be considered as the independent piece of information (see also the discussion of the spatial resolution estimation problem below).

The meaning of the columns and lines of AKM is straightforward:

- The column of AKM with index ‘*j*’ shows how the variation of an individual element of vector $\boldsymbol{x}$ with index ‘*j*’ will be mapped to the part of solution delivered by measurements that are described by the operator **K** (it should be stressed that another part of the solution is delivered by *a priori* information in the form of the mean profile).

- The line of AKM with index '*j*' shows how the variations of individual elements of vector $\boldsymbol{x}$ will contribute to the value of element '*j*' of the part of the solution delivered by measurements that are described by the operator **K**.

It is possible to obtain a rough estimate of the spatial resolution from this mapping in two ways. The first way is to estimate the half width at half maximum of the function represented by the matrix column that corresponds to the desired coordinate (desired altitude in specific case). The description of another way to obtain the vertical resolution of the retrievals can be found in the paper by Schneider et al. (2008). The following quantity can be taken as the value of the vertical resolution at given altitude z_j:

$$\Delta z_j = \min(z_{j+n(j)} - z_{j-n(j)}) \text{ in case } h_j \geq 1, \text{ where } h_j = \sum_{l=j-n(j)}^{j+n(j)} \mathbf{A}_{ll}, \tag{10}$$

where $n(j)$ is the number of neighbouring matrix elements providing the validity of the condition $h_j \geq 1$. So, according to Schneider et al. (2008), the altitude range can be identified as an independent layer for the retrieval if the sum of the corresponding diagonal elements of the averaging kernel matrix reaches unity.

The total altitude region of the retrieval can be estimated from the analysis of the quantity

$$g_p = \sum_{l=1}^{L} \mathbf{A}_{pl}, \tag{10a}$$

where '*p*' and '*l*' are matrix element indices. In the papers by Vigouroux et al. (2008) and Lindenmaier et al. (2010), the term 'sensitivity' of the retrievals to the measurements is used for this quantity, which indicates, at each altitude z_p, the fraction of the retrieval that comes from the measurement rather than from the *a priori* information. Different criteria can be used for the estimation of reliable retrievals: in the paper by Vigouroux et al. (2008) the value $g_p > 0.5$ has been taken.

Before proceeding to the description of the general approach, it is important to make several remarks. First, if we look at the cost function (Equation (2)), we can notice that mathematically there is no difference between measurements and *a priori* information. In other words, the algorithm 'does not know' which term is relevant to *a priori* information and which term is relevant to actual information. Both terms in the expression for the cost function are of equal importance. Second, as we can notice, the *a priori* information in case of optimal estimation can be treated as *in situ* virtual measurements of the elements of vector $\boldsymbol{x}$; in this case, the matrix operator of measurements is equal to the unity operator **I**, and the cost function can be presented as

$$\begin{aligned} f_c = &(\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0))^{\mathrm{T}} \mathbf{S}^{-1} (\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0)) \\ &+ (\boldsymbol{r} - \boldsymbol{r}_0 - \mathbf{I}(\boldsymbol{x} - \boldsymbol{x}_0))^{\mathrm{T}} \mathbf{D}^{-1} (\boldsymbol{r} - \boldsymbol{r}_0 - \mathbf{I}(\boldsymbol{x} - \boldsymbol{x}_0)), \end{aligned} \tag{11}$$

where **D** is the covariance matrix of the sought parameter and $\boldsymbol{r}$ corresponds to the results of virtual *in situ* measurements and in our particular case, when linearization is related to statistical mean values, the following relation is valid: $\boldsymbol{r} = \boldsymbol{r}_0 = \boldsymbol{x}_0$. If we use an arbitrary initial value $\boldsymbol{x}_0$ for linearization of the problem, then $\boldsymbol{r}$ would be equal to statistical mean and $\boldsymbol{r} \neq \boldsymbol{r}_0 = \boldsymbol{x}_0$.

Apparently, the Tikhonov regularization can be treated as virtual measurements of the derivative of the sought function with respect to the spatial coordinate and one can write:

$$f_c = (\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0))^{\mathrm{T}}\mathbf{S}^{-1}(\boldsymbol{y} - \boldsymbol{y}_0 - \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_0)) + (\boldsymbol{r} - \boldsymbol{r}_0 - \mathbf{L}(\boldsymbol{x} - \boldsymbol{x}_0))^{\mathrm{T}}\alpha\mathbf{I}(\boldsymbol{r} - \boldsymbol{r}_0 - \mathbf{L}(\boldsymbol{x} - \boldsymbol{x}_0)), \tag{12}$$

where $\boldsymbol{r}$ corresponds to the results of *in situ* measurements of the derivative of the sought function, $\boldsymbol{r}_0$ corresponds to the values of the derivative of the initial state $\boldsymbol{r}_0 = \mathbf{L}\boldsymbol{x}_0$, and for Tikhonov regularization $\boldsymbol{r} = \boldsymbol{r}_0$. Again, an arbitrary value could be chosen for $\boldsymbol{x}_0$, and then we would have $\boldsymbol{r} \neq \boldsymbol{r}_0$.

These remarks lead us to the conclusion that we can use the unified description for both actual measurements and *a priori* information and derive the unified formulae for any multi-parameter and multi-constrained problem. Before we pass to the formulation of the general approach, a small note concerning terminology should be made. In the paper by Rodgers (1976), the term 'virtual measurements' has been introduced for *a priori* information and the term 'direct measurements' has been used for conventional measurements. In the book by Rodgers (2000), the term 'real measurements' has been used for conventional measurements. In the present article, we prefer to use the term 'actual' instead of 'direct' or 'real' since this term reflects more precisely the main difference between *a priori* information and conventional measurements.

2.2. *The general approach*

The general approach to the formulation and solution of the inverse problem consists of the same steps as traditional approach but has the following principal features:

(1) The vector of sought parameters is composed of target and interfering parameters.
(2) Linearization of the radiative transfer equation is done with respect to target parameters and interfering parameters as well. Several different types of measurements, *a priori* information, and constraints are taken into account simultaneously.
(3) Measurements and constraints are incorporated into the algorithm in the same form: constraints are expressed as virtual measurements related to the sought vector by means of linearized operators.
(4) The reference values of sought parameters that are used for linearization do not necessarily correspond to statistical mean values.
(5) The errors of different types of measurements are not correlated.

Let us consider these features in more detail. As an example, we introduce the following vector of sought parameters:

$$\boldsymbol{x}^{\mathrm{T}} = (a_1, a_2, \ldots, a_L,\ b_1, b_2, \ldots, b_L,\ c_1, c_2, \ldots, c_L,\ d_1, d_2, \ldots, d_L), \tag{13}$$

where a, b, c, and d are parameters describing the state and (or) composition of the atmosphere; indices '1, 2, ..., L' denote spatial coordinates (altitude levels); and superscript 'T' denotes transposition. At the moment it is not important for us to determine which parameters are target ones and which are interfering ones. It should be noted that such situations may exist wherein the sub-dimension of one or several parameters is not equal to 'L' – for example, in case when the coefficients of parameterizations are considered as unknown values. We also note that

- there are no limitations for including non-atmospheric parameters (e.g. instrument parameters) into sought vector; and
- the components of the $\boldsymbol{x}$ vector are not necessarily sensitive to all components of the measurements.

In order to illustrate the concept of actual and virtual measurements, we present Figure 1. The term 'actual measurements' refers in general case to different remote and *in situ* measurements that provide information about target and interfering parameters at specific moment or period of time. The set of these measurements therefore corresponds to a kind of synergistic measurement scheme. The information that cannot be referred exactly to specific moment or period of time of actual measurements is classified as 'virtual measurements' or '*a priori* information'. *A priori* information under general approach may comprise different type of data including statistics, constraints, model relationships, and physical relationships. Any combination of actual and virtual measurements can be used; however, one should keep in mind that some combinations may not deliver the solution since we consider the actual remote measurements for which the correspondent inverse problem is ill-posed.

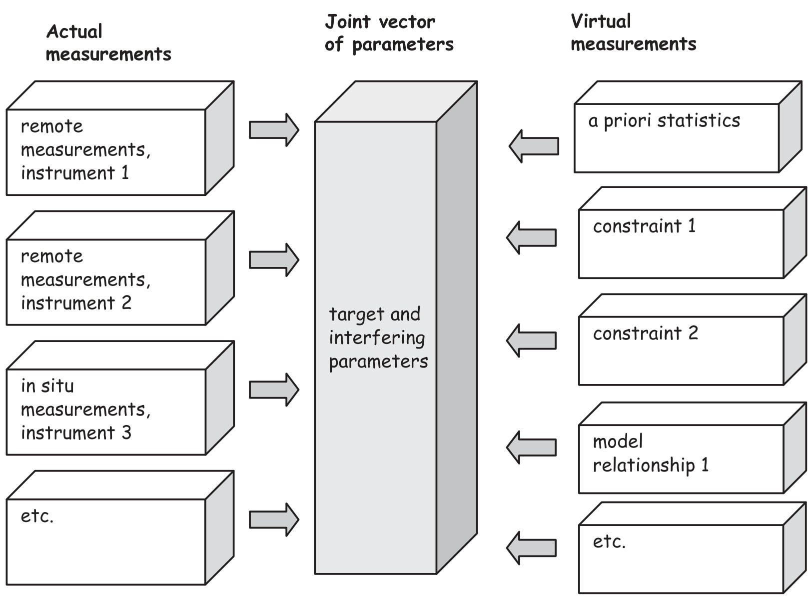

Figure 1. The block chart illustrating the concept of actual and virtual measurements under the general approach to the formulation of the inverse problems.

The system of equations to be solved under the general approach will be the following:

$$\begin{cases} \boldsymbol{y}_1 - \boldsymbol{y}_{10} = \mathbf{K}_1(\boldsymbol{x} - \boldsymbol{x}_0) \\ \boldsymbol{y}_2 - \boldsymbol{y}_{20} = \mathbf{K}_2(\boldsymbol{x} - \boldsymbol{x}_0) \\ \vdots \quad \vdots \quad \vdots \\ \boldsymbol{y}_N - \boldsymbol{y}_{N0} = \mathbf{K}_N(\boldsymbol{x} - \boldsymbol{x}_0) \end{cases}, \tag{14}$$

where N is the total number of actual and virtual measurements, $\boldsymbol{y}_i$ and $\mathbf{K}_i$ are the vectors of the results of actual and virtual measurements and corresponding linear operators, and the zero index denotes the reference values for linearization. The cost function of the least-square method can be written now as

$$f_\mathrm{c} = \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{y}_{i0} - \mathbf{K}_i(\boldsymbol{x} - \boldsymbol{x}_0))^\mathrm{T} \mathbf{S}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{y}_{i0} - \mathbf{K}_i(\boldsymbol{x} - \boldsymbol{x}_0)), \tag{15}$$

where $\mathbf{S}_i$ are the error matrices corresponding to actual and virtual measurements. It should be stressed that the presentation of the cost function as the sum of a number of terms corresponding to specific type of measurements is possible since we assumed the errors of measurements of different types being not correlated. One can notice that it is possible to combine all results of all measurements into one mega-vector and to combine all operators into one mega-operator. We do not use such formulation since

- it is more convenient for analysis to use separate matrices; and
- we do not consider cases where the errors of measurements of different types are correlated.

In order to find the minimum of the cost function (15), we write the derivative of the scalar function f_c with respect to vector argument $\boldsymbol{x}$ and solve again Equation (5). We obtain

$$\boldsymbol{x} = \boldsymbol{x}_0 + \left(\sum_{i=1}^{N} \mathbf{K}_i^\mathrm{T} \mathbf{S}_i^{-1} \mathbf{K}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{K}_i^\mathrm{T} \mathbf{S}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{y}_{i0}) \right). \tag{16}$$

The expression for the error matrix (the Fisher information matrix) of the solution is the following:

$$\mathbf{F} = \left(\sum_{i=1}^{N} \mathbf{K}_i^\mathrm{T} \mathbf{S}_i^{-1} \mathbf{K}_i \right)^{-1}. \tag{17}$$

Now we have the solution (Equation (16)) and the error estimation (Equation (17)). Let us consider the averaging kernel matrix in the general case for the purpose of deriving the degrees of freedom (i.e. DOFS) and the spatial resolution. For the sake of convenience, let us separate measurements designated by index 'J' and other measurements with indices $j \neq J$. Using Equation (17) we rewrite Equation (16):

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_0 + \sum_{j=1}^{N} \left(\mathbf{F}\mathbf{K}_j^{\mathrm{T}} \mathbf{S}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{y}_{j0}) \right) \\ &= \boldsymbol{x}_0 + \sum_{j \neq J} \left(\mathbf{F}\mathbf{K}_j^{\mathrm{T}} \mathbf{S}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{y}_{j0}) \right) + \mathbf{F}\mathbf{K}_J^{\mathrm{T}} \mathbf{S}_J^{-1} (\boldsymbol{y}_J - \boldsymbol{y}_{J0}) \\ &= \boldsymbol{x}_{\mathrm{c}} + \mathbf{G}_J (\boldsymbol{y}_J - \boldsymbol{y}_{J0}), \end{aligned} \tag{18}$$

where

$$\mathbf{G}_J = \mathbf{F}\mathbf{K}_J^{\mathrm{T}} \mathbf{S}_J^{-1} \tag{18a}$$

can be called the solving operator (gain matrix) corresponding to specific type of measurements designated by index 'J', and the vector $\boldsymbol{x}_{\mathrm{c}}$ is the reference vector plus contributions to the solution from measurements for which $j \neq J$. Let us replace $(\boldsymbol{y}_J - \boldsymbol{y}_{J0})$ with $\mathbf{K}_J(\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0)$, where $\boldsymbol{x}_{\mathrm{t}}$ denotes the true value of the sought vector. We obtain

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_{\mathrm{c}} + \mathbf{G}_J (\boldsymbol{y}_J - \boldsymbol{y}_{J0}) \\ &= \boldsymbol{x}_{\mathrm{c}} + \mathbf{G}_J \mathbf{K}_J (\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0) \\ &= \boldsymbol{x}_{\mathrm{c}} + \mathbf{A}_J (\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0), \end{aligned} \tag{19}$$

where

$$\mathbf{A}_J = \mathbf{G}_J \mathbf{K}_J = \mathbf{F}\mathbf{K}_J^{\mathrm{T}} \mathbf{S}_J^{-1} \mathbf{K}_J \tag{19a}$$

is the averaging kernel matrix corresponding to measurements of the type designated by index 'J'. The physical meaning of the AKM is the same as its meaning under the traditional approach: the columns of the AKM show how the variation of individual element of vector $\boldsymbol{x}$ (the state parameter local variation equal to unity) will be mapped to the part of solution delivered by measurements of the type 'J'.

Under the traditional approach, AKMs are often used in the validation schemes when it is necessary to compare the results of the remote measurements with independent information on the atmospheric state that are made with high spatial resolution and taken as a 'true solution'. In order to fulfil this task, it is necessary to provide AKM and the reference vector $\boldsymbol{x}_0$. In general case one more vector is needed: $\boldsymbol{x}_{\mathrm{c}}$ which represents the reference vector plus contributions to the solution from the measurements other than measurements for which AKM is provided. We note that this vector is not needed under the traditional approach just because it coincides with the reference vector. In the general case one should distinguish actual and virtual measurements before providing AKMs and vectors for the purpose of validation and comparison. All virtual measurements that describe *a priori* information of different types and provide constant contribution to the solution should be taken into account in calculations of the $\boldsymbol{x}_{\mathrm{c}}$ vector:

$$\boldsymbol{x}_{\mathrm{c}} = \boldsymbol{x}_0 + \sum_{\text{virtual}} \left(\mathbf{F}\mathbf{K}_j^{\mathrm{T}} \mathbf{S}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{y}_{j0}) \right). \tag{20}$$

For all actual measurements, the combined AKM should be calculated:

$$\mathbf{A}_{\text{actual}} = \sum_{\text{actual}} \left(\mathbf{F}\mathbf{K}_j^{\mathrm{T}} \mathbf{S}_j^{-1} \mathbf{K}_j \right). \tag{21}$$

Now, if the 'true' state vector is available, its smoothed variant can be calculated and used for comparison:

$$\boldsymbol{x} = \boldsymbol{x}_{\mathrm{c}} + \mathbf{A}_{\text{actual}} (\boldsymbol{x}_{\mathrm{t}} - \boldsymbol{x}_0). \tag{22}$$

It is easy to show and is very important to stress that

$$\sum_{j=1}^{N} \mathbf{A}_j = \mathbf{I}, \tag{23}$$

and so

$$\sum_{j=1}^{N} \mathrm{tr}(\mathbf{A}_j) = (\mathrm{DOFS})_{\text{total}}, \tag{24}$$

i.e. the sum of AKMs for all considered types of measurements is equal to the unity matrix and the sum of traces of AKMs is equal to the total number of elements of the sought vector $\boldsymbol{x}$, which, accordingly, is equal to the maximal DOFS of the problem. When analysing DOFS, one should keep in mind the complicated structure of the AKM:

$$\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{aa} & \mathbf{A}_j^{ab} & \mathbf{A}_j^{ac} & \mathbf{A}_j^{ad} \\ \mathbf{A}_j^{ba} & \mathbf{A}_j^{bb} & \mathbf{A}_j^{bc} & \mathbf{A}_j^{bd} \\ \mathbf{A}_j^{ca} & \mathbf{A}_j^{cb} & \mathbf{A}_j^{cc} & \mathbf{A}_j^{cd} \\ \mathbf{A}_j^{da} & \mathbf{A}_j^{db} & \mathbf{A}_j^{dc} & \mathbf{A}_j^{dd} \end{pmatrix}, \tag{25}$$

where index j refers to the measurement type and different matrix blocks correspond to different parameters of atmospheric state and composition (*a, b, c, d*) as stated in the definition of the sought vector (Equation (13)). As it follows from Equations (24) and (25), the DOFS for different parameters can be analysed separately. So, it could be useful, for example, to estimate the contribution of each type of measurements to the derivation of each parameter in terms of DOFS in the form of Table 1.

Let us consider the spatial resolution. Similar to the traditional approach, the following quantity can be taken as the value of the spatial (vertical) resolution provided by measurements of certain type at a given coordinate (altitude) z_j:

$$\Delta z_j = \min(z_{j+n(j)} - z_{j-n(j)}) \ \text{ in case } \ h_j \geq 1, \ \text{where} \ h_j = \sum_{l=j-n(j)}^{j+n(j)} \mathbf{A}_{1\ ll}^{aa}, \tag{26}$$

where we consider measurements of type '1' and the parameter '*a*' as an example, and '*ll*' denotes the indices of diagonal elements of the matrix $\mathbf{A}_1$. If the sum of the diagonal elements of the averaging kernel matrix corresponding to sequence of neighbouring coordinates reaches unity, the spatial (altitude) range determined by these coordinates

Table 1. The contributions of actual and virtual measurements to the derivation of parameters expressed in terms of DOFS values.

	Measurements				
	Actual		Virtual		
Parameter	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
a	$\mathrm{tr}\left(\mathbf{K}_1^{aa}\right)$	$\mathrm{tr}\left(\mathbf{K}_2^{aa}\right)$	$\mathrm{tr}\left(\mathbf{K}_3^{aa}\right)$	$\mathrm{tr}\left(\mathbf{K}_4^{aa}\right)$	$\mathrm{tr}\left(\mathbf{K}_5^{aa}\right)$
b	$\mathrm{tr}\left(\mathbf{K}_1^{bb}\right)$	$\mathrm{tr}\left(\mathbf{K}_2^{bb}\right)$	…	…	…
c	$\mathrm{tr}\left(\mathbf{K}_1^{cc}\right)$	…	…	…	…
d	…	…	…	…	…

can be identified as an independent layer for the retrieval of given parameter using given type of measurements.

Similar to Equation (10a), the altitude region of the retrieval can be estimated from the analysis of the quantity

$$g_{j_p} = \sum_{l=1}^{M} \mathbf{A}_{j_{pl}}, \tag{27}$$

where 'j' is the index denoting measurement type and 'p' and 'l' are matrix element indices; summation is over column number from 1 to M, where M is the total number of elements in the combined vector $\boldsymbol{x}$. The quantity g in Equation (27) shows the fraction of the retrieval of the element of combined sought vector $\boldsymbol{x}$ with index 'p' that comes from the measurement with index 'j' rather than from measurements of all other types. It is important to stress two circumstances:

(1) The matrix blocks (Equation (25)) may have different dimensions, if the combined sought vector comprises different physical quantities.
(2) The natural variability of physical quantities, which form the sought vector may differ considerably from each other and also may be different at different spatial coordinates.

Therefore, in order to avoid misleading values of g it would be more correct in physical sense to rewrite Equation (27) as

$$g_{j_p} = \frac{1}{\boldsymbol{v}_p} \sum_{l=1}^{L} \mathbf{A}_{j_{pl}} \boldsymbol{v}_l, \tag{28}$$

where vector $\boldsymbol{v}$ is composed of typical variations of sought parameters and the dimensions of corresponding elements of vectors $\boldsymbol{v}$ and $\boldsymbol{x}$ are identical. One can easily show that for each 'p',

$$\sum_{j=1}^{N} g_{j_p} = 1. \tag{29}$$

The quantities presented in Table 1 and in Equations (26)–(29) refer to measurements of one specific type, which can be called 'individual AKMs'. In practice, it is useful to calculate these quantities for a group of measurements of different types. The most evident situation is the analysis of DOFS and spatial resolution provided by all actual measurements – in this case, we use the combined AKM given by Equation (21) instead of the 'individual' AKMs. It should be noted that any combination of individual actual and virtual measurements can be used for estimations in the process of analysis of inverse problem under the general approach. So, we introduce the AKM of a group of measurements of selected types:

$$\mathbf{A}_{\text{group}} = \sum_{\text{group}} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j \right). \tag{30}$$

For such group of measurements, one can calculate DOFS, the spatial resolution, and the spatial region of retrieval using $\mathbf{A}_{\text{group}}$ as was described above for the measurements of one single type.

Now, having introduced a group of measurements, we can return to the error matrix and consider its decomposition similar to presentation of the total error under traditional approach as the sum of smoothing error and random noise (see Equation (7a)). The error matrix under the general approach (Equation (17)) can be expressed in the form of sum of contributions from each type of measurements:

$$\mathbf{F} = \sum_{j=1}^{N} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j\mathbf{F} \right). \tag{31}$$

If measurements are combined in two groups as actual and virtual, then we have

$$\mathbf{F} = \sum_{\text{actual}} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j\mathbf{F} \right) + \sum_{\text{virtual}} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j\mathbf{F} \right). \tag{32}$$

In this case, using the terminology of Rodgers (2000), the first term can be considered as the 'random noise' and the second term can be considered as the 'smoothing error'. And finally, if we use arbitrary division of measurements into groups, then we can express the error matrix as

$$\mathbf{F} = \sum_{\text{group1}} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j\mathbf{F} \right) + \sum_{\text{group2}} \left(\mathbf{F}\mathbf{K}_j^{\text{T}}\mathbf{S}_j^{-1}\mathbf{K}_j\mathbf{F} \right) + \cdots, \tag{33}$$

where the total number of terms in all summations is equal to the total number of measurement types.

In order to minimize the errors caused by the non-linearity of the original operators describing actual and virtual measurements, the solution can be obtained in the iteration process based on Gauss–Newton method. The cost function in this case will be as follows:

$$f_{\text{c}} = \sum_{i=1}^{N} \left(\boldsymbol{y}_i - \boldsymbol{y}_{i,k} - \mathbf{K}_{i,k}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \right)^{\text{T}} \mathbf{S}_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{y}_{i,k} - \mathbf{K}_{i,k}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \right), \tag{34}$$

where k is the iteration step number, $\boldsymbol{y}_{i,k}$ is the vector of measurements of type 'i' calculated with the help of corresponding non-linearized original operator, and $\mathbf{K}_{i,k}$ is the linearized operator calculated using the solution $\boldsymbol{x}_k$. In order to find the minimum of the cost function (34), we write the derivative of the scalar function f_c with respect to vector argument $\boldsymbol{x}_{k+1}$ and solve Equation (5) again. We obtain

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \left(\sum_{i=1}^{N} \mathbf{K}_{i,k}^{\mathrm{T}} \mathbf{S}_i^{-1} \mathbf{K}_{i,k}\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{K}_{i,k}^{\mathrm{T}} \mathbf{S}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{y}_{i,k})\right). \tag{35}$$

There is a completely equivalent formula in which the iteration step is expressed relative to any vector $\boldsymbol{x}_0$ taken as a reference:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_0 + \left(\sum_{i=1}^{N} \mathbf{K}_{i,k}^{\mathrm{T}} \mathbf{S}_i^{-1} \mathbf{K}_{i,k}\right)^{-1} \left[\sum_{i=1}^{N} \mathbf{K}_{i,k}^{\mathrm{T}} \mathbf{S}_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{y}_{i,k} + \mathbf{K}_{i,k}(\boldsymbol{x}_k - \boldsymbol{x}_0)\right)\right]. \tag{36}$$

It should be stressed that $\boldsymbol{x}_0$ is not necessarily an '*a priori* state vector' since under the general approach there is no distinguished '*a priori* state vector'. If it is used in one of the constraints, it is transformed into the result of one of the virtual measurements. Equation (34) is written in the general form as well as subsequent Equations (35) and (36).

3. Examples of presenting *a priori* information and constraints as virtual measurements

3.1. Statistical mean and covariances

Using statistical mean and covariances as *a priori* information corresponds to the well-known optimal estimation method. As it was indicated in Section 2.1, the statistical mean is taken as virtual measurement of the sought vector, the measurement operator is equal to unity operator and the errors are described by statistical covariance matrix. If we consider for simplicity the case with just one type of actual measurements and one optimal estimation type constraint, then we shall have

$$\boldsymbol{y}_2 = \boldsymbol{x}_0, \tag{37}$$

$$\mathbf{K}_2 = \mathbf{I}, \tag{37a}$$

$$\mathbf{S}_2 = \mathbf{D}, \tag{37b}$$

where $\mathbf{D}$ is the joint covariance matrix (with cross-covariances) for the complete set of parameters, which in our case comprises the parameters $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$.

3.2. Spatial gradients

Using information on spatial gradients as *a priori* information corresponds to the well-known Tikhonov regularization method. As it was indicated in Section 2.1, the desired spatial gradient is taken as virtual measurement, and the measurement operator is the differential quadrature operator **L**. Again, if we consider for simplicity just one type of actual measurements and one Tikhonov-type constraint, then we shall have

$$\boldsymbol{y}_2 = \mathrm{d}\boldsymbol{x}_0/\mathrm{d}s, \tag{38}$$

$$\mathbf{K}_2 = \mathbf{L}. \tag{38a}$$

Where s is a spatial coordinate. The error matrix of the virtual measurements of spatial gradients is subject to fitting in every specific case similar to fitting of the parameter 'α' in the classic Tikhonov regularization method. Kulawik et al. (2006) introduced a method for developing a constraint matrix based on altitude-varying combinations of zeroth-, first-, and second-order derivatives of an unknown parameter profile.

3.3. *Iteration shift constraint*

In order to obtain fast and stable convergence of the iteration process (Equations (35) and (36)) in case of strong non-linearity of the measurement operators, it is useful to put a limit to a solution shift at several initial iteration steps. The iteration shift constraint is equivalent to the standard Levenberg–Marquardt method. This method and the questions relevant to convergence of an iteration process are discussed in detail in the book (Rodgers 2000). Here, we only note that in this case we 'virtually measure' the sought vector at each iteration step and obtain the following 'result of virtual measurements':

$$\boldsymbol{y} = \boldsymbol{x}_k. \tag{39}$$

The operator of these virtual measurements is the unity operator and the error matrix is subject to fitting in order to provide stable convergence. A very important notice should be made: among other virtual measurements, the iteration shift virtual measurement is the 'most virtual', since it has no physical background. Therefore, this type of virtual measurements must be excluded from the process of calculation of DOFS, error matrix, spatial resolution, and retrieval ranges.

3.4. *Numerical model predictions and assigning spatial boundaries for retrievals*

The incorporation of numerical model predictions into the retrieval algorithm is the same as for statistical mean and covariances. The error matrix should correspond to the accuracy of model predictions of sought parameters. Assigning spatial boundaries for retrievals is equivalent to high-accuracy *in situ* virtual measurements of the parameter in the areas where this parameter is assumed to be known. It should be noted that depending on a character and temporal resolution of these constraints, they may be attributed either to actual or to virtual measurements.

3.5. *Water vapour: transfer from absolute humidity to relative humidity*

In case the amount of water vapour is incorporated in the forward and inverse problems in the form of a number density profile or an absolute humidity profile, sometimes it could be natural to apply constraints to a relative humidity profile. This could be done if we use the relationship

$$r = f(q, T), \tag{40}$$

where r and q are relative and absolute humidity, and T is the temperature, f is a function converting absolute humidity into relative humidity. If T belongs to a set of unknowns, after linearization we have

$$\delta r = \frac{\partial f}{\partial q}\delta q + \frac{\partial f}{\partial T}\delta T. \tag{41}$$

The results of virtual measurements will be the values of relative humidity and the corresponding error matrix will be the covariance matrix for relative humidity.

4. Application of the general approach to the problem of sounding the non-LTE atmosphere

4.1. The breakdown of the local thermodynamic equilibrium

At certain atmospheric altitudes and at certain atmospheric conditions, the assumption of the local thermodynamic equilibrium (LTE) cannot be used for calculations of the radiative transfer in the IR spectral bands of a number of atmospheric gases. Due to a series of collisional, radiative, and chemical processes, the populations of the vibrational states (and in some cases of the rotational states) of molecules deviate from Boltzmann Law. As an example, one can mention the violation of LTE (non-LTE conditions) in the middle and upper atmosphere in the ro-vibrational bands of carbon dioxide and ozone. The altitude levels of the LTE breakdown depend on the specific molecule and the vibrational state and, for example, can be as low as 30 km for the high-energy vibrational states of ozone molecules (Manuilova et al. 1998). For CO_2 states that give origin to 15 μm band, the non-LTE effect starts at about 70 km altitude (Edwards, Lopez-Puertas, and Lopez-Valverde 1993). Under non-LTE conditions, populations of the vibrational states of molecules are often described by means of the so-called vibrational temperatures T_v (the values of T_v are used in the standard Boltzmann expression for population instead of the values of kinetic temperature T_k, which is used under the LTE conditions). The difference between vibrational temperatures and kinetic temperature characterizes the magnitude of the 'non-LTE effect'. One can find the comprehensive state-of-the-art study of the non-LTE effect in the book by Lopez-Puertas and Taylor (2001).

4.2. Methods of sounding the non-LTE atmosphere

The investigation of the temperature regime and gas composition of the middle atmosphere (the mesosphere and the lower thermosphere in particular) by means of the space-borne remote sensing in the infrared region using limb geometry is carried out intensively due to several reasons. First, large number of modern satellite instruments is sensitive enough to register weak emissions of the atmospheric constituents at high altitudes. Second, high spectral resolution of spectrometers and interferometers gives the possibility to extract specific information that is contained in spectral signatures of different gases and absorption bands. Third, the well-known advantages of the space-borne remote-sensing technique such as global coverage and the possibility for long-term observations on regular basis should also be mentioned.

Two main approaches to the problem of sounding the non-LTE atmosphere can be mentioned. The main idea of the first approach is the derivation of the information on the non-equilibrium populations of the vibrational states of molecules directly from spectra simultaneously with the kinetic temperature and gas concentration (Zachor and Sharma 1985; Timofeyev, Kostsov, and Grassl. 1995). The second approach is based on the application of numerical models that describe processes driving non-LTE to the retrieval algorithm (López-Puertas et al. 1998; Zhou et al. 1998). Both approaches have advantages and disadvantages as well as specific application areas. The second approach can be applied to the measurements with medium and low spectral resolution or with high resolution but with limited number of spectral channels, when the extraction of information on the population of distinct vibrational states from spectral signatures is problematic. In this respect, the instruments ISAMS (Improved Stratospheric And Mesospheric Sounder), SABER (Sounding of the Atmosphere using Broadband Emission Radiometry), and LIMS (Limb Infrared Monitor of the Stratosphere) can be mentioned. In this case, the estimation of the non-LTE effect on the basis of preliminary modelling is the only possible solution. However, preliminary modelling requires the precise knowledge of physics and quantitative characteristics of population processes, e.g. their rate constants, but the relevant information is insufficient in many cases and is characterized by large uncertainties.

The main advantages of the first approach are (1) no necessity to apply population models that cannot be completely adequate and (2) the possibility to derive simultaneously the information about many parameters, which leads to the possibility to improve the population models on the basis of retrieved data. However, the application of this method requires measurements with sufficiently high spectral resolution since the extraction of information on the population of specific vibrational states of different gases and their isotopic variants is possible only when corresponding spectral signatures are resolved. It should be mentioned that due to the large number of unknowns (e.g. the inverse problem formulated in Timofeyev, Kostsov, and Grassl (1995) that comprised more than 30 unknown profiles – kinetic temperature, pressure, CO_2 and O_3 abundances, vibrational temperatures of 11 vibrational states for CO_2 and 18 states for ozone), the multi-parameter retrieval technique accounting for non-LTE consumes a lot of computing time and therefore special efforts are needed for its adaptation to operational use.

4.3. The CRISTA experiment (brief description)

The experiments with the CRISTA instrument were successfully performed in November 1994 (CRISTA-1) and August 1997 (CRISTA-2) on board the Shuttle Pallet Satellite (SPAS). Limb scan measurements of atmospheric infrared radiance were done in the 4–71 μm spectral region. Observations were performed in several modes including the high-altitude one with the tangent altitude range from 40 to 150 km. The accuracy of tangent altitude pointing was ±200 m. This mode is of particular interest for studying the state of the middle atmosphere and the non-LTE effect for different molecules. The detailed description of the experiments and the presentation of part of the results can be found in the papers by Offermann et al. (1999) and Riese et al. (1999). A review of the main scientific results from two CRISTA missions has been presented by Grossmann et al. (2004).

Processing of the high-altitude atmospheric scans were done independently by several scientific groups. The group from Wuppertal University used the second of the above-mentioned approaches to sounding the non-LTE atmosphere (with preliminary modelling of the non-equilibrium populations of the vibrational states). As a result, the global distributions of CO_2 in mesosphere and lower thermosphere have been obtained from 4.3 μm measurements and reported by Kaufmann et al. (2002). In the present article, we consider the application of the first of the above-mentioned approaches (developed at St Petersburg State University) to the interpretation of the limb radiance data obtained during CRISTA experiments in 1994 and 1997.

4.4. Input data and the formulation of the non-LTE inverse problem

Limb radiance measurements in the 15 μm CO_2 band and in the 9.6 μm O_3 band have been processed by the multi-parameter and multi-constrained method; however, in the present article, we describe the results corresponding to the 15 μm CO_2 band only. The key features of the input data are presented in Table 2. It should be noted that the spectral range of the spectrometer SCS8_ext is slightly shifted relative to the spectral range of the SCL2 spectrometer.

Table 2. Key features of the input data (limb radiance spectra in 15 μm CO_2 band).

Experiment	CRISTA-1	CRISTA-2
Time period	5–12 November 1994	8–16 August 1997
Spectrometer	SCL2	SCS8_ext
Spectral range, cm^{-1}	642–685	661–700
Tangent altitude range, km	40–150	40–140
Number of limb scans processed accounting for non-LTE	881	792
Spectral resolving power, $\lambda/\Delta\lambda$	525	525
Wavenumber step, cm^{-1}	0,25	0,25
Tangent altitude step, km	about 2.5 (variable)	about 2.5 (variable)

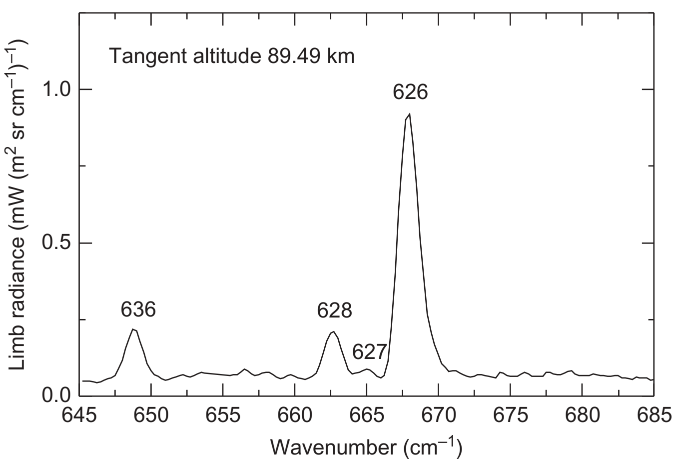

Figure 2. Sample spectrum (CRISTA-1 experiment). Numbers near the spectral signatures identify the regions of the Q-branches of the transitions between states 01101, 02201, and 03301 corresponding to four isotopic variants of the CO_2 molecules considered in the study (626, 636, 628, 627).

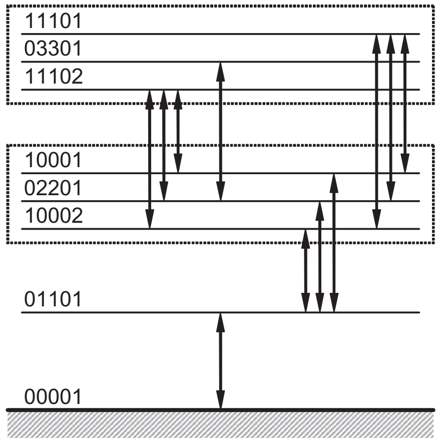

Figure 3. The vibrational transitions and states of the CO_2 molecules for which the non-LTE conditions were taken into account. The boxes marked by dashed lines indicate the groups of levels with the so-called internal LTE.

As an example, one of the recorded spectra is shown in Figure 2. The vibrational transitions and the states of the CO_2 molecules for which the non-LTE conditions were taken into account are shown in Figure 3 (the 'internal LTE' groups of levels are discussed below).

The sought vector was composed of 31 parameters at 43 altitude levels (from 35 to 160 km with steps of 2.5 km): pressure, kinetic temperature, CO_2 number density, and seven vibrational temperatures of the states, 01101, 10002, 02201, 10001, 11102, 03301, 11101, of four most abundant isotopic variants of CO_2 molecules (626, 636, 628, and 627). The relation between registered limb radiance and sought parameters can be written in the form of a functional:

$$I(v, z_t) = A\left(T_k(z), p(z), n(z), T_v^{g,s}(z)\right), \tag{42}$$

where I is the limb radiance at wavenumber v and tangent altitude z_t, A denotes non-linear forward operator, z is the vertical coordinate, p is pressure, T_k is kinetic temperature, n is CO_2 number density, T_v denotes vibrational temperatures, 'g' is the index identifying CO_2 isotopic variant, and 's' is the index identifying vibrational state.

So, the problem is to retrieve the vertical distributions of all sought parameters from the spectral measurements. The physical basis for such retrieval should be explained. Apparently, the vertical distribution can be obtained from limb scanning, since a spectrum registered at specific tangent altitude contains information on the atmospheric state mainly in the layers above this tangent altitude. So the main problem is to 'separate the variables' at a given altitude. Since gas content is taken in the form of the number density rather than volume mixing ratio (VMR), pressure affects the radiation intensity in the radiative transfer equation through the Lorentz line-shape only. Therefore, in the layers where collisional broadening is predominant, pressure can be derived from spectral shape. In the higher layers, where Doppler broadening is predominant, pressure can be retrieved on the basis of the application of the hydrostatic equilibrium constraint to the temperature profile and the pressure values in the lower layers. The separation of the other variables (kinetic temperature, gas number density, and vibrational

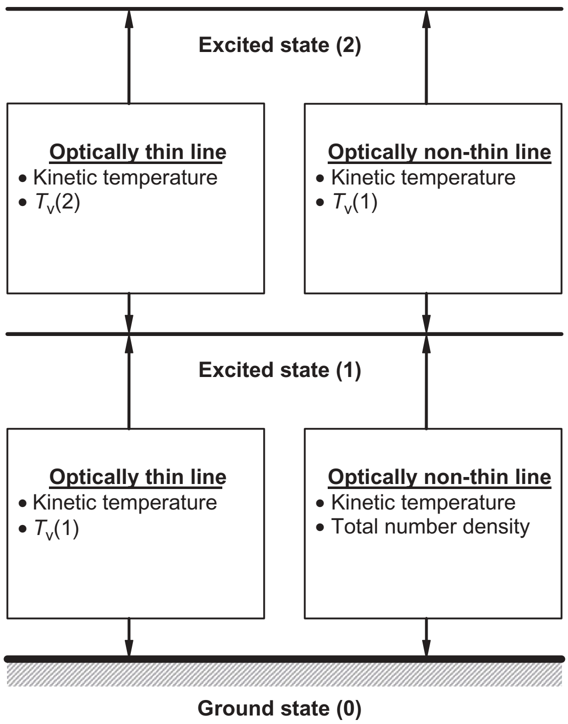

Figure 4. The diagram illustrating 'separation of the variables' in the problem of multi-parameter retrieval accounting for the non-LTE conditions.

temperatures) is illustrated in Figure 4. We consider, for simplicity, the transitions between ground (0) and two excited (1 and 2) vibrational states. Let us divide spectral lines into two classes – optically thin and optically non-thin. For the case with optically thin lines, the radiation intensity is determined mainly by the emission, so it contains the information on the population of the upper state. For the case with optically non-thin lines, the radiation intensity is determined both by the populations of the upper and lower states (both emission and absorption make their contribution). Hence, the vibrational temperature T_{v2} can be uniquely determined from the measurements in the optically thin line of the transition 2–1. Then, knowing T_{v2} (population of the state 2), the vibrational temperature of the state 1 can be obtained from the measurements in the non-thin line of the transition 2–1. The same procedure works also for the fundamental transition (1–0) lines. It should be noted that the derivation of the population of the ground state practically is equivalent to the derivation of the total number density. It should also be stressed that measurements in the optically thin line of the fundamental transition are complementary to the measurements in the optically non-thin line of transition 2–1 and also deliver the information on the population of the state 1 (T_{v1}). The kinetic temperature enters the radiative transfer equation through rotational substructure of spectra (rotational states are assumed to be in LTE) and through Doppler line shape. Therefore, measurements of radiation intensity in different branches of the vibrational bands give the possibility to extract independent information also on the kinetic temperature. Taking into account the fact that the bands consist of lines with intensities that differ by several orders of magnitude, in real situation, when radiance is measured with high and medium spectral resolution, we have the possibility to utilize the information on the considered variables from different lines at different altitudes.

4.5. Description of constraints

Despite the fact that there is the physical basis for the 'separation of the variables', because of overlapping of spectral lines and finite spectral resolution of measurements the considered inverse problem remains ill-posed and should be constrained. The expression for the cost function corresponding to the non-LTE inverse problem is the same as given by formula (15). The actual measurements are the limb radiation measurements made by the CRISTA instrument in the CO_2 absorption bands. They are described by the linearized radiative transfer operator. The rest ($N - 1$) of the measurements are virtual and correspond to constraints that are presented in detail below.

4.5.1. Transfer from number density to mixing ratio

If the amount of atmospheric gas is incorporated in the forward and inverse problems in the form of a number density profile, sometimes it could be more convenient to apply constraints to a VMR profile. (This is our case since there is much *a priori* information on the CO_2 VMR.) It can be easily done if virtual measurements of VMR are assumed with corresponding measurement operator based on the relationship

$$v = n\frac{\mathrm{k_B}T}{p}, \tag{43}$$

where v and n are VMR and number density values, T is temperature, p is pressure, and $\mathrm{k_B}$ is Boltzmann's constant. The error matrix of these virtual measurements will be equivalent to the covariance matrix of VMR. If T and p are included in the set of unknowns, then the linearized operator should be written with respect to all parameters in Equation (43) and it will be the following:

$$\delta v = \frac{kT}{p}\delta n + \frac{nk}{p}\delta T + \frac{-nkT}{p^2}\delta p. \tag{44}$$

4.5.2. Hydrostatic equilibrium

As has been stated above, the hydrostatic equilibrium condition is necessary for retrieving pressure in the layers where collisional broadening is not predominant. This condition can also be incorporated in the retrieval procedure by means of virtual measurements approach. The operator of virtual measurements is based on the equation relating the pressure values p at neighbouring altitude levels z_l and z_{l+1}:

$$p(z_{l+1}) = p(z_l)\exp\left(-\left(\frac{\mu_l g_l}{RT_l} + \frac{\mu_{l+1} g_{l+1}}{RT_{l+1}}\right)\frac{z_{l+1} - z_l}{2}\right), \tag{45}$$

where μ is the molecular mass of air, g is the acceleration of gravity, R is the gas constant, and T is temperature. In order to introduce the virtual measurements that describe the considered relationship, we rewrite Equation (45) as follows:

$$y_{l+1} = -p(z_{l+1}) + p(z_l)\exp\left(-\left(\frac{\mu_l g_l}{RT_l} + \frac{\mu_{l+1} g_{l+1}}{RT_{l+1}}\right)\frac{z_{l+1} - z_l}{2}\right), \tag{46}$$

where $y_{l+1} = 0$ ($l = 1, L - 1$) are considered as virtual measurements. The physical meaning of the error of virtual measurements in the presented example is the expected difference between the retrieved pressure value at level $l + 1$ and the value that satisfies the hydrostatic equilibrium equation exactly. So, in the considered example, we have $L - 1$ virtual measurements and corresponding $(L - 1) \times (L - 1)$ error matrix with all elements equal to zero except diagonal elements. If both pressure and temperature belong to a set of unknowns, the expression (46) should be linearized accordingly.

It should be emphasized that the incorporation of hydrostatic equilibrium constraint into the retrieval scheme in the manner described above is especially important for limb-viewing instruments due to the following reason: subsequent tangent altitudes are shifted horizontally in one limb scan due to the orbital motion of the instrument (rear-looking and downward scanning compensate this shift only partly). As a result, the hydrostatic equilibrium approximation should work well for neighbouring pairs of altitudes, but not necessarily should be valid for the whole profile.

4.5.3. *The LTE breakdown altitude level*

Since the vibrational temperatures are included in the set of unknowns, the methodology of virtual measurements gives a good possibility to take into account various situations relevant to physics of non-LTE processes. For example, it is known that up to definite altitudes the LTE assumption is valid. In our case with 15 μm CO_2 band, this altitude constitutes about 60–65 km. This situation can be expressed in terms of virtual measurements of the difference between vibrational and kinetic temperature:

$$y = T_v - T_k. \tag{47}$$

In the altitude range where LTE is valid, $y = 0$ with high accuracy.

4.5.4. *Internal LTE*

Another example is the so-called internal LTE effect, which means that the populations of the closely located (with respect to their energy) vibrational states relate to each other according to Boltzmann law (an example is presented in Figure 3). This effect for two vibrational states can be described by the relationship

$$\frac{E_{v2} - E_{v1}}{T_k} = \frac{E_{v2}}{T_{v2}} - \frac{E_{v1}}{T_{v1}}, \tag{48}$$

where E_v is the energy of the vibrational state, with the indices '1' and '2' referring to the lower and upper states, respectively. The expression for the virtual measurements is

$$y = -\frac{E_{v2} - E_{v1}}{T_k} + \frac{E_{v2}}{T_{v2}} - \frac{E_{v1}}{T_{v1}}. \tag{49}$$

Evidently, if 'internal LTE' is valid, $y = 0$ with high accuracy.

4.5.5. Excited state constraint

One more constraint is stipulated by the characteristics of the instrument and the experiment. The CRISTA instrument has moderate spectral resolution, which is about 1.2 cm^{-1} in the considered spectral interval. This spectral resolution is not sufficient to separate in spectra the Q-branches (the rotational quantum number in the ground state is the same as the rotational quantum number in the excited state) corresponding to the transitions between the states 01101, 02201, and 03302 for all considered isotopic variants of the CO_2 molecules. For each isotopic variant, these Q-branches are presented in spectra as one single signature but not as three distinguishable signatures (Figure 2). As a result, separation of the variables $T_{v(01101)}$, $T_{v(02201)}$, and $T_{v(03301)}$ is problematic. The numerical experiments have shown that there is a possibility to get independent information on these variables which is contained in the shape of the signatures. However, due to the fact that the instrument line-shape function is known with limited accuracy, in practice such possibility does not exist. The uncertainty of the instrument line shape can cause artefacts in the retrieved profiles of the vibrational temperatures. In order to stabilize the solution, the following constraint has been applied, which will be referred to below as 'excited state' constraint:

$$T_{v(01101)} = T_{v(02201)} = T_{v(03301)}. \quad (50)$$

In the retrieval procedure, this constraint is incorporated in the form of virtual measurements of two differences between vibrational temperatures:

$$y = T_{v(01101)} - T_{v(02201)}, \quad (51)$$

$$y = T_{v(01101)} - T_{v(03301)}, \quad (52)$$

where $y = 0$ with high accuracy.

Taking into account the 'internal LTE' constraint and the 'excited state' constraint, only four parameters corresponding to the non-LTE populations remain unknown. These parameters can be called 'effective' vibrational temperatures for four most abundant isotopes: T_{v626}, T_{v636}, T_{v628}, and T_{v627}. So, due to the application of multiple constraints, instead of 31 unknown parameters now we have only seven: kinetic temperature, pressure, CO_2 VMR, and four effective vibrational temperatures for CO_2 isotopic variants, 626, 636, 628, and 627.

Table 3 shows the main constraints that have been used for processing limb radiance measurements in the 15 μm CO_2 band obtained in the CRISTA space experiment. As can be seen from Table 3, there is no parameter that is not coupled with another by one or more physical constraints. Kinetic temperature is coupled with all parameters, pressure – with kinetic temperature and CO_2 number density. Higher than 65 km, where LTE is no more valid, there are groups of vibrational temperatures influenced by internal LTE constraint. And the vibrational temperature of the state 01101 is coupled with vibrational temperatures of the states 02201 and 03301 in the whole altitude range. (It should be noted that not all constraints are shown in the table, and that there have been additional constraints applied individually to each parameter.) As a result of application of multiple constraints, the ill-posed inverse problem is well regularized and the set of retrieved parameters is self-consistent.

Table 3. The main constraints that have been used for processing of the limb radiance measurements in the 15 μm CO_2 band (the CRISTA space experiment) by the multi-parameter algorithm.

		Main constraints					
Parameters		Hydrostatic equilibrium	*A priori* information on CO_2 VMR	LTE up to 65 km	Internal LTE higher than 65 km		Excited state constraint
p (pressure)		X	X				
T_k (kinetic temperature)		X	X	X	X	X	
CO_2 total number density			X				
T_v for CO_2 626 isotopic variant	T_v 01101			X			X
	T_v 10002			X	X		
	T_v 02201			X	X		X
	T_v 10001			X	X		
	T_v 11102			X		X	
	T_v 03301			X		X	X
	T_v 11101			X		X	
T_v for CO_2 636 isotopic variant: constraints are the same as for 626 isotopic variant							
T_v for CO_2 628 isotopic variant: constraints are the same as for 626 isotopic variant							
T_v for CO_2 627 isotopic variant: constraints are the same as for 626 isotopic variant							

Note: The symbol 'X' denotes that a corresponding parameter has been influenced by a corresponding constraint.

4.6. Results: the CO_2 vibrational temperatures

Part of the results of the interpretation of the CRISTA high-altitude limb radiances by the multi-parameter and multi-constrained method accounting for non-LTE conditions has already been published. In the paper by Kostsov, Timofeyev, and Manuilova (2004) and in a series of papers by Kostsov and Timofeev (2003, 2005, 2005a, 2005b), the mesospheric temperature inversions and the global distributions of CO_2 and ozone in the mesosphere and lower thermosphere have been analysed. In this section, we present some results of the retrievals of the CO_2 vibrational temperatures from the data of CRISTA-1 and CRISTA-2 experiments, which have not yet been published.

It has been shown in the previous section that the multi-parameter method provided the possibility to retrieve the four so-called effective vibrational temperatures. The special numerical study has shown that for three less abundant isotopic variants of the CO_2 molecules (636, 628, and 627) the effective vibrational temperatures are practically equal to the vibrational temperature of the 01101 state in the whole altitude range considered. For the most abundant isotopic variant (626), the effective vibrational temperature is very close to the vibrational temperature of the 02201 state up to 90 km altitude, and above 90 km, it is practically equal to the vibrational temperature of the 01101 state. Below, keeping this fact in mind, we shall in most cases omit the word 'effective'.

The retrieval errors for the vibrational temperature profiles are given in Table 4. It should be stressed that the upper boundaries of the retrievals are different because of different intensities of spectral lines, which provide the information on the specific vibrational temperatures.

A qualitative comparison of the results of the retrieval of vibrational temperatures in the CRISTA experiment with model predictions has been done using non-LTE models that have been developed in the course of preparations for the MIPAS space experiment by López-Puertas et al. (1998). Mean profiles of the kinetic and vibrational temperatures corresponding to summer conditions in polar latitudes are presented in Figure 5. One important remark should be made: the 15 μm band spectra registered during the CRISTA-2 experiment and processed in this study do not contain the range of the Q-branches of the bands of the 636 isotopic variant. Consequently, the vibrational

Table 4. The random (ran), systematic (sys), and total (tot) retrieval errors for the vibrational temperature profiles of the four isotopic variants of the CO_2 molecules (626, 636, 628, 627).

	T_{v626} (K)			T_{v636} (K)			T_{v628} (K)			T_{v627} (K)		
z (km)	ran	sys	tot	ran	sys	tot	ran	sys	tot	ran	sys	tot
40–65	The altitude range of the 'LTE up to 65 km' constraint.											
70	1.0	0.5	1.1	1.1	0.7	1.3	1.1	0.7	1.3	1.4	0.9	1.7
75	1.2	0.5	1.3	1.3	0.6	1.5	1.4	0.7	1.5	1.9	1.1	2.2
80	1.5	0.4	1.6	1.7	0.6	1.8	1.8	1.5	2.3	2.8	2.1	3.5
85	1.8	0.4	1.8	2.2	1.1	2.5	2.6	3.8	4.6	3.9	4.6	6.0
90	2.1	0.4	2.1	3.8	3.5	5.2	4.5	5.4	7.0	6.4	5.3	8.3
95	2.5	0.4	2.5	5.8	7.4	9.4	8.3	7.3	11.1	–	–	–
100	3.0	1.0	3.1	7.6	11.7	13.9	9.9	10.4	14.4	–	–	–
105	4.0	4.9	6.3	10.0	14.4	17.6	–	–	–	–	–	–
110	7.2	10.6	12.8	13.1	15.6	20.3	–	–	–			
115	16.6	13.3	21.2	–		–						
120	18.3	13.4	22.7									

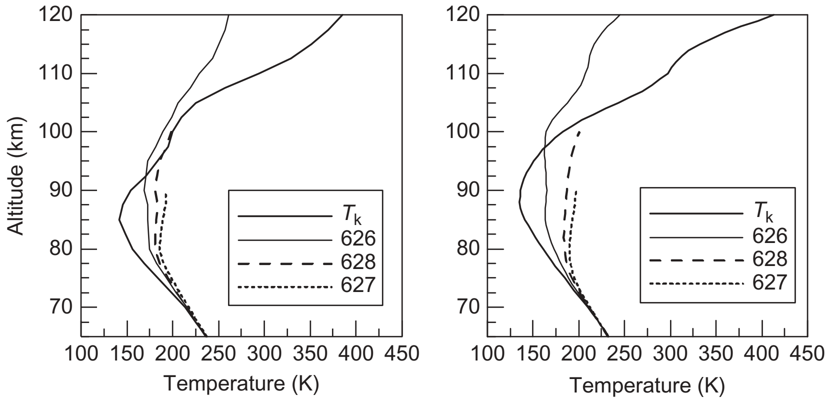

Figure 5. The vibrational temperatures for three isotopic variants of the CO_2 molecules (626, 628, 627) as obtained from CRISTA-2 data (left) and from model calculations (López-Puertas et al. 1998) (right). Polar latitudes, summer.

temperature $T_{v(636)}$ was not retrieved in the CRISTA-2 experiment. It is well known that the kinetic temperature profile in polar latitudes in summer has cold and low mesopause. The mesopause altitudes corresponding to the experimental and model data are 85 and 88 km. The obtained profiles of vibrational temperatures are in good agreement with the model profiles. The deviation of the vibrational temperatures from the kinetic temperature starts at 70 km altitude and the magnitude of the non-LTE effect is practically the same. The non-LTE effect in the mesopause region for 626 isotopic variant of the CO_2 molecules is 30 K both for the experiment and for the model, the non-LTE effects for 628 and 627 isotopic variants are 40 and 47 K for the experiment while the model prediction gives 50 and 60 K.

The comparison of the experimental and the model profiles of the vibrational temperatures for polar latitudes and winter conditions is presented in Figure 6. Under these conditions, the kinetic temperature profiles are characterized by high warm mesopause: 100 km (CRISTA-2) and 97 km (model). The non-LTE effect for all isotopic variants of

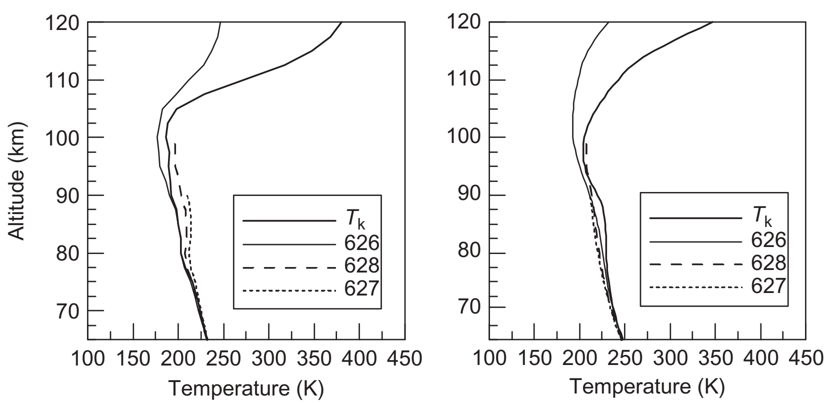

Figure 6. Vibrational temperatures for three isotopic variants of the CO_2 molecules (626, 628, 627) as obtained from CRISTA-2 data (left) and from model calculations (López-Puertas et al. 1998) (right). Polar latitudes, winter.

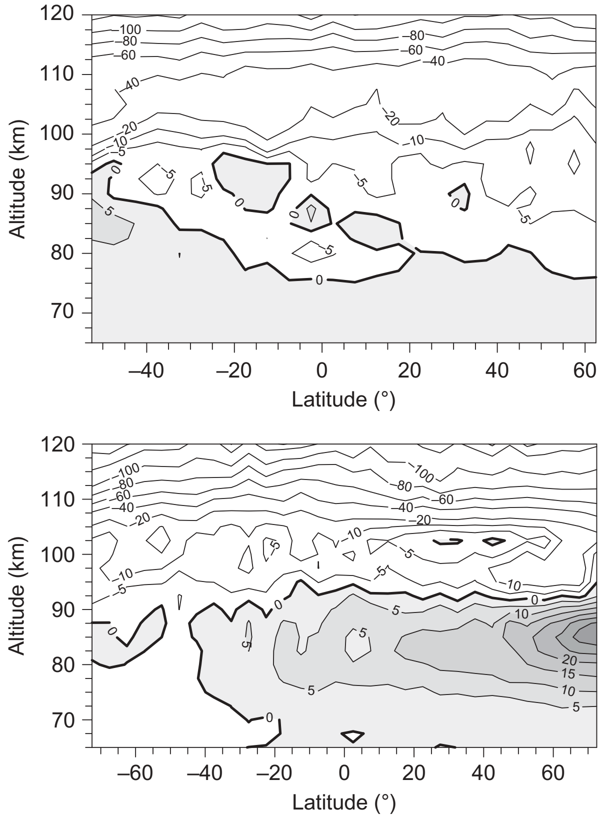

Figure 7. Zonal-mean values of the non-LTE effect ($\Delta = T_v - T_k$, Kelvin) for the isotopic variant 626 of the CO_2 molecule as obtained from CRISTA-1 (top) and CRISTA-2 (bottom) data.

the CO_2 molecules is not so considerable, if compared to summer conditions in polar latitudes. The level of the LTE breakdown is about 75 km for both the experimental and model profiles.

Zonal mean values of the non-LTE effect obtained from the CRISTA-1 and CRISTA-2 data are presented in Figures 7–9. The CRISTA-1 experiment was carried out in November, so the data corresponded to late autumn conditions in the Northern Hemisphere and late spring conditions in the Southern Hemisphere. The CRISTA-2 experiment was carried out in August and the obtained data corresponded to purely summer conditions in the Northern Hemisphere and purely winter conditions in the Southern Hemisphere. As a result, the latitudinal differences of the non-LTE effect are more pronounced in the CRISTA-2 data. When analysing the results, one should also pay attention to the fact that latitudinal coverage during the second CRISTA experiment was larger than during the first one. One can see that for all isotopic variants of the CO_2 molecules, the non-LTE effect is maximal in the polar regions in astronomical summer. This phenomena is explained by the excitation of vibrational states due to upward and downward heat fluxes coming from warmer regions below and above the mesopause. In winter, when the temperature gradient in the mesosphere is not strong, the non-LTE effect is much smaller; however, it can be considerable for less abundant CO_2 isotopic variants. The presented experimentally obtained zonal mean values of the non-LTE effect are in

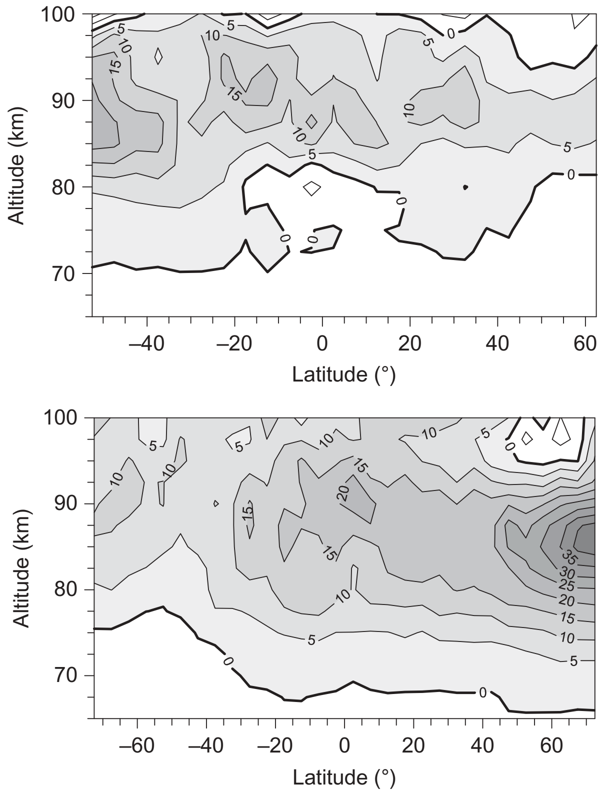

Figure 8. Zonal-mean values of the non-LTE effect ($\Delta = T_v - T_k$, Kelvin) for the isotopic variant 628 of the CO_2 molecules as obtained from CRISTA-1 (top) and CRISTA-2 (bottom) data.

good agreement with model predictions published earlier by Lopez-Puertas, Lopez-Valverde, and Taylor (1992).

5. Conclusion

In the present article, the unified approach and the corresponding formulae are presented and discussed, which can be used for solving multi-parameter ill-posed inverse problems of atmospheric remote sensing based on combination of measurements of different type and multiple constraints. The approach is based on the concept of virtual measurements proposed by Rodgers (1976) and his idea that many of the apparently different methods of solving inverse problems are particular cases of the same general method.

The key points of the approach are as follows:

- *A priori* information of different types (statistics, constraints, model predictions, physical links between parameters, etc.) is considered and expressed as virtual measurements of finite accuracy.
- The error matrices of virtual measurements are subject to fitting in each specific case.

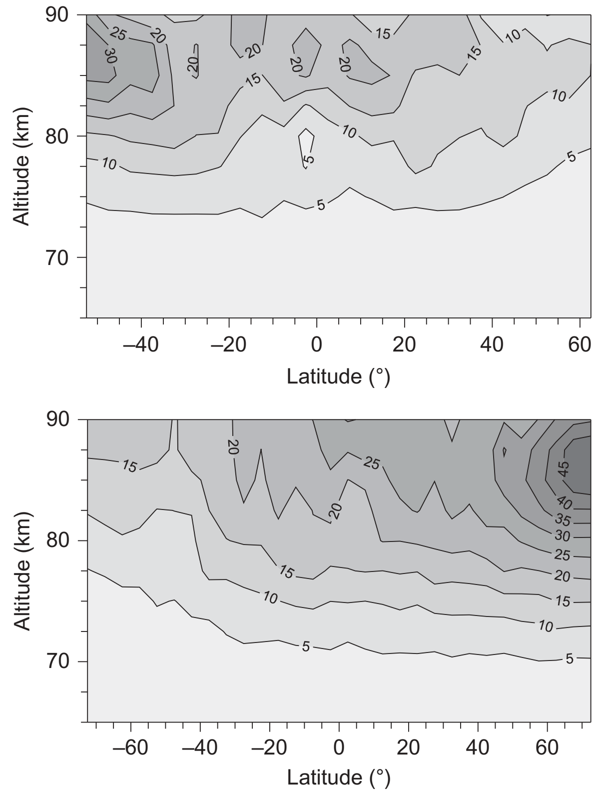

Figure 9. Zonal-mean values of the non-LTE effect ($\Delta = T_v - T_k$, Kelvin) for the isotopic variant 627 of the CO_2 molecule as obtained from CRISTA-1 (top) and CRISTA-2 (bottom) data.

- The term 'actual measurements' refers to different remote and *in situ* measurements that provide the information about target and interfering parameters at specific moment or period of time and that are the core measurements of a remote-sensing experiment.
- Mathematically, there is no difference between actual and virtual measurements.
- Measurements of different types can be combined into groups, and for each group the information on corresponding retrieval errors, DOFS, averaging kernels, and spatial resolution can be obtained.
- Any combination of actual and virtual measurements can be used in the retrieval procedure; however, one should keep in mind that some combinations may not deliver the solution since the inverse problem corresponding to actual remote measurements is assumed to be ill-posed.

Examples of different constraints expressed as virtual measurements are presented, which demonstrate the way how various physical relationships can be incorporated into retrieval algorithms.

The peculiar features and the results of the practical application of the general approach to processing the limb radiance data obtained during the CRISTA experiments have been described.

Acknowledgements

The author is grateful to K. Grossmann, D. Offermann, M. Kaufmann, J. Oberheide, and other members of the CRISTA experiment team for the access to CRISTA limb radiance data and for the cooperation.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by Russian Science Foundation [Project No. 14-17-00096].

ORCID

V.S. Kostsov http://orcid.org/0000-0002-6160-6634

References

Aires, F. 2011. "Measure and Exploitation of Multisensor and Multiwavelength Synergy for Remote Sensing: 1. Theoretical Considerations." *Journal of Geophysical Research* 116: D02301. doi:10.1029/2010JD014701.

Clarmann, T. 2014. "Smoothing Error Pitfalls." *Atmospheric Measurement Techniques* 7: 3023–3034. doi:10.5194/amt-7-3023-2014.

Clarmann, T., U. Grabowski, and M. Kiefer. 2001. "On the Role of Non-Random Errors in Inverse Problems in Radiative Transfer and Other Applications." *Journal of Quantitative Spectroscopy & Radiative Transfer* 71: 39–46. doi:10.1016/S0022-4073(01)00010-3.

Doicu, A., F. Schreier, S. Hilgers, A. von Bargen, S. Slijkhuis, M. Hess, and B. Aberle. 2007. "An Efficient Inversion Algorithm for Atmospheric Remote Sensing with Application to UV Limb Observations." *Journal of Quantitative Spectroscopy & Radiative Transfer* 103: 193–208. doi:10.1016/j.jqsrt.2006.05.007.

Edwards, D. P., M. Lopez-Puertas, and M. Lopez-Valverde. 1993. "Non-Local Thermodynamic Equilibrium Studies of the 15-µm Bands of CO_2 for Atmospheric Remote Sensing." *Journal of Geophysical Research* 98: 14955–14977.

Grossmann, K. U., O. Gusev, M. Kaufmann, A. Kutepov, and P. Knieling. 2004. "A Review of the Scientific Results from the CRISTA Missions." *Advances in Space Research* 34: 1715–1721. doi:10.1016/j.asr.2003.02.041.

Kaufmann, M., O. A. Gusev, K. U. Grossmann, R. Roble, M. Hagan, C. Hartsough, and A. A. Kutepov. 2002. "The Vertical and Horizontal Distribution of CO_2 Densities in the Upper Mesosphere and Lower Thermosphere as Measured by CRISTA." *Journal of Geophysical Research* 107: 8182. doi:10.1029/2001JD000704.

Koner, P. K., and J. R. Drummond. 2008. "A Comparison of Regularization Techniques for Atmospheric Trace Gases Retrievals." *Journal of Quantitative Spectroscopy & Radiative Transfer* 109: 514–526. doi:10.1016/j.jqsrt.2007.07.018.

Kostsov, V. S., and Y. M. Timofeev. 2005. "Mesospheric Temperature Inversions from CRISTA-1 Data." *Izvestiya Atmospheric and Oceanic Physics* 41: 741–749.

Kostsov, V. S., and Y. M. Timofeev. 2005a. "Mesospheric Ozone from the CRISTA-1 Satellite Experimental Data: 1. Method of Profile Determination and its Accuracy." *Izvestiya Atmospheric and Oceanic Physics* 41: 178–190.

Kostsov, V. S., and Y. M. Timofeev. 2005b. "Mesospheric Ozone from the CRISTA-1 Satellite Experimental Data: 2. Spatial Distributions and Diurnal Variations." *Izvestiya Atmospheric and Oceanic Physics* 41: 191–202.

Kostsov, V. S., and Y. M. Timofeyev. 2003. “Mesospheric Carbon Dioxide Content as Determined from the CRISTA-1 Experimental Data.” *Izvestiya Atmospheric and Oceanic Physics* 39: 322–332.

Kostsov, V. S., Y. M. Timofeyev, and R. O. Manuilova. 2004. “Global Distributions of Temperature, Carbon Dioxide, Ozone, and Non-LTE Parameters in Mesosphere and Lower Thermosphere (CRISTA-1 Experiment).” *Proceedings of SPIE* (5235): 208–219. doi:10.1117/12.512269.

Kulawik, S. S., G. Osterman, D. B. A. Jones, and K. W. Bowman. 2006. “Calculation of Altitude-Dependent Tikhonov Constraints for TES Nadir Retrievals.” *IEEE Transactions on Geoscience and Remote Sensing* 44 (5): 1334–1342. doi:10.1109/TGRS.2006.871206.

Lindenmaier, R., R. L. Batchelor, K. Strong, H. Fast, F. Goutail, F. Kolonjari, C. T. McElroy, R. L. Mittermeier, and K. A. Walker. 2010. “An Evaluation of Infrared Microwindows for Ozone Retrievals Using the Eureka Bruker 125HR Fourier Transform Spectrometer.” *Journal of Quantitative Spectroscopy & Radiative Transfer* 111: 569–585. doi:10.1016/j.jqsrt.2009.10.013.

Lopez-Puertas, M., M. A. Lopez-Valverde, and F. W. Taylor. 1992. “Vibrational Temperatures and Radiative Cooling of the CO_2 15 μm Bands in the Middle Atmosphere.” *Quarterly Journal of the Royal Meteorological Society* 118: 499–532.

Lopez-Puertas, M., and F. W. Taylor. 2001. “Non-LTE Radiative Transfer in the Atmosphere.” In *Series on Atmospheric, Oceanic and Planetary Physics*, edited by F. W. Taylor, Vol. 3, 487. Singapore: World Scientific.

López-Puertas, M., G. Zaragoza, M. Á. López-Valverde, F. J. Martín-Torres, G. M. Shved, R. O. Manuilova, A. A. Kutepov, O. Gusev, T. V. Clarmann, A. Linden, G. Stiller, A. Wegner, H. Oelhaf, D. P. Edwards, and J.-M. Flaud. 1998. “Non-Local Thermodynamic Equilibrium Limb Radiances for the MIPAS Instrument on Envisat-1.” *Journal of Quantitative Spectroscopy & Radiative Transfer* 59: 377–403. doi:10.1016/S0022-4073(97)00128-3.

Manuilova, R. O., O. Gusev, A. A. Kutepov, T. V. Clarmann, H. Oelhaf, G. P. Stiller, A. Wegner, M. López-Puertas, F. J. Martín-Torres, G. Zaragoza, and J.-M. Flaud. 1998. “Modelling of Non-LTE Limb Spectra of i.r. Ozone Bands for the MIPAS Space Experiment.” *Journal of Quantitative Spectroscopy & Radiative Transfer* 59: 405–422. doi:10.1016/S0022-4073(97)00120-9.

Offermann, D., K. U. Grossmann, P. Barthol, P. Knieling, M. Riese, and R. Trant. 1999. “The Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) Experiment and Middle Atmosphere Variability.” *Journal of Geophysical Research* 104: 16311–16325.

Riese, M., R. Spang, P. Preusse, M. Ern, M. Jarisch, D. Offermann, and K. Grossmann. 1999. “Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) Data Processing and Atmospheric Temperature and Trace Gas Retrieval.” *Journal of Geophysical Research* 104: 16349–16367.

Rodgers, C. D. 1976. “Retrieval of Atmospheric Temperature and Composition from Remote Measurements of Thermal Radiation.” *Reviews of Geophysics and Space Physics* 14 (4): 609–624.

Rodgers, C. D. 1998. “Information Content and Optimisation of High Spectral Resolution Remote Measurements.” *Advances in Space Research* 21 (3): 361–367. doi:10.1016/S0273-1177(97)00915-0.

Rodgers, C. D. 2000. “Inverse Methods for Atmospheric Sounding: Theory and Practice.” In *Series on Atmospheric, Oceanic and Planetary Physics*, edited by F. W. Taylor, Vol. 2, 238. Singapore: World Scientific.

Rodgers, C. D., and B. J. Connor. 2003. “Intercomparison of Remote Sounding Instruments.” *Journal of Geophysical Research* 108 (D3): 4116. doi:10.1029/2002JD002299.

Schneider, M., F. Hase, T. Blumenstock, A. Redondas, and E. Cuevas. 2008. “Quality Assessment of O_3 Profiles Measured by a State-of-the-Art Ground-Based FTIR Observing System.” *Atmospheric Chemistry and Physics* 8: 5579–5588. doi:10.5194/acp-8-5579-2008.

Steck, T., and T. v. Clarmann. 2001. “Constrained Profile Retrieval Applied to the Observation Mode of the Michelson Interferometer for Passive Atmospheric Sounding.” *Applied Optics* 40 (21): 3559–3571.

Timofeyev, Y. M., V. S. Kostsov, and H. Grassl. 1995. “Numerical Investigations of the Accuracy of the Remote Sensing of Non-LTE Atmosphere by Space-Borne Spectral Measurements of Limb I.R. Radiation: 15 μm CO_2 Bands, 9.6 μm O_3 Bands and 10 μm CO_2 Laser Bands.” *Journal of*

Quantitative Spectroscopy & Radiative Transfer 53: 613–632. doi:10.1016/0022-4073(95)00025-G.

Vigouroux, C., M. DeMazière, P. Demoulin, C. Servais, F. Hase, T. Blumenstock, I. Kramer, M. Schneider, J. Mellqvist, A. Strandberg, V. Velazco, J. Notholt, R. Sussmann, W. Stremme, A. Rockmann, T. Gardiner, M. Coleman, and P. Woods. 2008. "Evaluation of Tropospheric and Stratospheric Ozone Trends over Western Europe from Ground-Based FTIR Network Observations." *Atmospheric Chemistry and Physics* 8: 6865–6886. doi:10.5194/acp-8-6865-2008.

Zachor, A. S., and R. D. Sharma. 1985. "Retrieval of Non-LTE Vertical Structure from a Spectrally Resolved Infrared Limb Radiance Profile." *Journal of Geophysical Research* 90: 467–475.

Zhou, D. K., M. G. Mlynczak, G. E. Bingham, J. O. Wise, and R. M. Nadile. 1998. "CIRRIS-1A Limb Spectral Measurements of Mesospheric 9.6-μm Airglow and Ozone." *Geophysical Research Letters* 25: 643–646. doi:10.1029/98GL00236.